Research Article | Vol. 14, No. 10 / 1 Oct 2023 / Biomedical Optics Express 5466
Biomedical Optics EXPRESS

Unobtrusive biometric data de-identification of fundus images using latent space disentanglement

ZHIHAO ZHAO,[1,2] SHAHROOZ FAGHIHROOHI,[1] JUNJIE YANG,[1,2] KAI HUANG,[3] NASSIR NAVAB,[1] MATHIAS MAIER,[2] AND M. ALI NASSERI[2,*]

[1]*TUM School of Computation, Information and Technology , Technical University of Munich, Arcisstrasse 21, Munich, 80333, Germany*
[2]*Klinik und Poliklinik für Augenheilkunde, Technische Universität München, IsmaningerStr. 22, München, 81675, Germany*
[3]*School of Computer Science and Engineering, Sun Yat-Sen University, Panyu District 132, Guangzhou, 510006, China*
[*]*nasseri@in.tum.de*

Abstract: With the incremental popularity of ophthalmic imaging techniques, anonymization of the clinical image datasets is becoming a critical issue, especially the fundus images, which would have unique patient-specific biometric content. Towards achieving a framework to anonymize ophthalmic images, we propose an image-specific de-identification method on the vascular structure of retinal fundus images while preserving important clinical features such as hard exudates. Our method calculates the contribution of latent code in latent space to the vascular structure by computing the gradient map of the generated image with respect to latent space and then by computing the overlap between the vascular mask and the gradient map. The proposed method is designed to specifically target and effectively manipulate the latent code with the highest contribution score in vascular structures. Extensive experimental results show that our proposed method is competitive with other state-of-the-art approaches in terms of identity similarity and lesion similarity, respectively. Additionally, our approach allows for a better balance between identity similarity and lesion similarity, thus ensuring optimal performance in a trade-off manner.

1. Introduction

Many organizations now publicly release large amounts of medical data in order to promote scientific research due to the growth of big data healthcare and retinal fundus photographs are also widely used in public presentations or reports as examples for the clinical screening and diagnosis of eye diseases [1,2]. Meanwhile, all contemporary health governing agencies and standards including ISO 22600-1:2014, The US Health Insurance Portability and Accountability Act (HIPAA) and European Data Protection Act incorporated measures to address patient privacy and clinical photographs. As per HIPAA regulations, the medical record can only be considered de-identified if any identifiable images are completely removed [3–5]. And Riis et al. also introduce a relatively new ethical transgression which is intrusion into a patient's bodily or social integrity by clinical descriptions and photographs printed in textbooks or medical journals [6]. Thereafter, the anonymization of the biometric information to prevent this leakage has become a critical issue. Although synthetic images are sometimes used by researchers in public presentations, they may not always provide the necessary details of specific pathologies that clinicians require. The most effective approach is to utilize real patient data while anonymizing the biometric information. The recognition of retinal biometric information is mainly based

#495438 https://doi.org/10.1364/BOE.495438
Journal © 2023 Received 15 May 2023; revised 29 Jul 2023; accepted 16 Aug 2023; published 26 Sep 2023

Biomedical Optics EXPRESS

on the vascular structure of retina images [7]. Although non-vascular features are also used as recognition features [8], the vascular structure of the retina remains relatively stable in one's lifespan [9]. This invariable stability and unforgettably has led researchers to primarily use vascular structures as identifying features of biological information [10,11]. As a result, *the best way is that we can use real patient data in pubic presentations or reports while anomymizing the biometric information by removing identification in vascular structure*.

There have been many different methods for de-identification. For example, image blurring, image masking, image inpainting and image obfuscation mentioned in [12–16]. However, these methods may lead to the loss of crucial information on pathologies. Meanwhile, these conventional methods turned out to be unreliable with the emergence of advanced algorithms like image denoising [17], occlusion region recovery [18] and visual reasoning [19]. To prevent attackers from inferring original data through the network, the differentially private and federated learning methods were proposed as other aspects of healthcare data privacy protection in [20–22]. Although these frameworks may prevent membership inference attacks, they cannot ensure the privacy of the original images. Some researchers use Generative Adversarial Networks(GANs) to generate plausible images instead of real data to protect privacy [23,24]. Conventional GAN models generate random images, limiting users' ability to edit those images to their requirements. To overcome this limitation, advanced de-identification GAN models now focus on disentanglement in latent space [25]. Therefore, *applying the disentanglement method for removing identification from fundus images without affecting pathologies needs to be the primary consideration*.

For the aforementioned problems, this paper proposes a method for fundus image de-identification on vascular structure while preserving the lesions such as hard exudates. In a well disentanglement latent space, one latent code can control the generation of a specific attribute [26]. The objective of this paper is to identify the latent code with highest contribution score to vascular structure in generated fundus image, as the retinal biometric identity features mainly exist in the retinal vascular structure [10,11]. As we want to preserve pathologies while removing identification, we use the hard exudate and vascular structure as our research object. Firstly, we train a fundus image generator based on StyleGAN3 [27]. In order to make the more realistic generated fundus images and better disentanglement for pathologies and vascular features, LPIPS [28]loss based on hard exudate detection and vessel segmentation are added into the network. After the generation model is well-trained, we propose a gradient-based method to determine the contribution score of each latent code to the vascular structure in the generated fundus image. To achieve this, we compute the gradient of the vessel in the fundus image for each latent code. Then we analyze the correlation between the gradient maps and the latent codes by calculating the sum of pixels in the gradient map. The position of latent code with highest contribution score to vascular structure in generated fundus image is determined according to the sum of pixels in gradient map. The major contributions of this paper can be included as follows:

1) A de-identification method is introduced by performing manipulation on the latent code with highest contribution score to identification while preserving pathologies.

2) Two LPIPS losses based on hard exudate segmentation and vessel segmentation are incorporated into StyleGAN3 for more realistic synthetic images and better disentanglement for pathologies and vascular features.

3) A gradient based method is proposed to locate the position of latent code that has the highest contribution score to identity features in fundus images.

2. Related work

Biometrics in Fundus Image. The stability, uniqueness, and non-duplicability of the vascular pattern make retinal recognition the most reliable biometric system [7]. The main elements of a

Biomedical Optics EXPRESS

fundus image captured by a digital fundus camera are the optic disc, blood vessels, and macula. Blood vessels are the most distinctive retinal feature used in personal identification [29]. The blood vessels in the retina create a branching pattern that looks like a tree, with the optic nerve head acting as the root on the surface of the retina [12]. The unique blood vessel pattern in the retina is used as a biometric identifier for personal identification. Changes in the rotation, movement, and size of the retinal image may undermine the reliability and safety of the biometric authentication process [9,30].

Local Editing with GAN. StyleGAN3 [27] has demonstrated a significant improvement in both image generation and disentanglement capabilities. Many current local image editing methods are based on StyleGAN structure [31–33]. The present research aims to discover a vector direction in the latent space using pretrained classifiers, which can be adjusted to manage specific attributes of an image, including age, gender, etc [34–36]. Meanwhile, StyleSpace [37] demonstrates that it is feasible to achieve precise control by adjusting the value of a single identified channel.

3. Methodology

Our paper presents a new method for local editing of the vascular structure region in fundus images by manipulating the meaningful latent code with the highest contribution score to the vessel. In this section, firstly, we introduce the training stage and LPIPS based loss function in section 3.1. Then we describe channel locating method to find the latent code with the highest contribution to vascular structure in section 3.2. Finally, we show how to apply de-identification on real image in section 3.3.

3.1. Training stage

3.1.1. Generation network structure

The network structure proposed in this paper is based on StyleGAN3. During the training phase of the network architecture, as depicted in Fig. 1, the network maps the input noise vector Z to the W space through a linear connection layer and then maps it to the S space via the affine transformation layer of StyleGAN3. Subsequently, the fundus image is generated by the decoder of StyleGAN3. In training process, while retaining the original stylegan loss function, we incorporate the LPIPS loss functions based on both vascular segmentation and lesion detection. During each training iteration, the depth features of both the generated and ground-truth image are extracted using the pre-trained vascular segmentation and lesion segmentation networks, respectively. The LPIPS loss is then calculated from these extracted deep features, and it can encourage the generator provides more realistic generated images and better disentanglement for pathologies and vascular features.

3.1.2. Loss functions

General loss function in GAN model is the adversarial loss L_{adv} which defined on the generator G and discriminator D of GAN:

$$\begin{aligned}\mathcal{L}_{adv} = &- \mathbb{E}_{I^{Gen},I^{GT}}[\log(D(I^{Gen}, I^{GT}))] \\ &- \mathbb{E}_{I^{Gen},I^{GT}}[\log(1 - D(I^{Gen}, I^{GT}))]\end{aligned} \tag{1}$$

where the target image I^{GT} comes from ground truth dataset, and synthetic image I^{Gen} comes from generator G. Since in this step we only need to implement the task of image generation, we only need to focus on the difference between the generated data distribution and the real data distribution, we don't need paired dataset in loss function calculating.

Due to the fundus dataset is not sufficiently large, only the adversarial loss can not get the desired performance. And inspired by the good performance of perceptual loss [38] in generation

Biomedical Optics EXPRESS

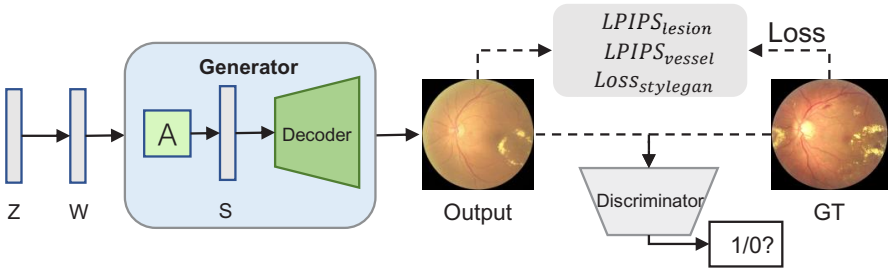

Fig. 1. The network structure consists of mapping layers that converts noise vector Z to W and S space, and a decoder structure that can generate fundus images from latent space. Additional loss functions are employed to encourage the generator provides more realistic generated images and better disentanglement for pathologies and vascular features.

tasks, we use LPIPS [28] to measure the deep perceptual features. General LPIPS is trained on 16-layer VGG network [39] pretrained on the ImageNet dataset [40], which is trained for image classification tasks and does not focus on retina fundus images. As we want to generate fundus images with more clear vascular structure and retina pathologies, we customized the LPIPS loss based on retina images tasks:

$$\mathcal{L}_{lpips} = \sum_{j} m^{j}(\phi^{j}(I^{Gen}) - \phi^{j}(I^{GT})) \tag{2}$$

where ϕ is the network we use to extract the deep perceptual features of target image I^{GT} and synthetic image I^{Gen}, m calculates the distance score of features and averages it from j layers.

In the paper, firstly, we train the vessel segmentation network based on U-Net [41]. Then we apply our vascular segmentation model to the generated images and randomly selected images from the test set, and select 8 feature layers respectively as the input for LPIPS, where 4 layers are the output after the downsampling layer of U-Net encoder, with the other 4 layers being the feature layers after the upsampling layer of decoder in U-Net. As such, the loss function will bring the vascular structure of generated images closer to that of real vessels. At the same time, to make the hard exudate more realistic, we apply the same method as in vascular segmentation, based on the trained U-Net model for lesion segmentation. Then, we employ cosine distance to measure the distance score. Finally, we add all loss functions together for training.

$$\mathcal{L}_{lpips} = LPIPS_{vessel}(I^{Gen}, I^{GT}) + LPIPS_{lesion}(I^{Gen}, I^{GT}) \tag{3}$$

$$\mathcal{L}oss = \lambda_{lpips}\mathcal{L}_{lpips} + \lambda_{adv}\mathcal{L}_{adv} \tag{4}$$

3.2. Locating latent code with highest contribution to vascular structure

Semantic manipulations can be perform by moving the latent code in the latent space along the direction of semantic areas. However, modifying the latent codes in the latent space is associated with spatial entanglements, which often lead to modifications of the image's irrelevant characteristics and features. To address this issue, recent researches on disentangled representations [42,43,35,37] believe that spatially disentangled manipulation of fine-grained controls can be achieved by modifying a single channel of the manipulation style. A latent representation is considered to be completely disentangled if each of its dimensions is responsible for one visual attribute. We explore the latent space as shown in Fig. 2. The projector that inverts real fundus image to latent space is provided by Karras et al. [27]. The generator is part of StyleGAN3 structure [27]. A pre-trained vessel segmentation network based on U-net [41] is applied to obtain the vascular mask. We compute the Jacobian matrix in masked region based on

Biomedical Optics EXPRESS

the vascular mask. The Jacobian matrix of each latent code is composed of the gradients in the forward propagation process, which means the contribution of the latent code to the vascular region [44]. Therefore, we analyze the correlation between the vascular structure and the latent codes by calculating the sum of pixels in each gradient map. The position of the latent code with the highest contribution score to the vascular structure in the generated fundus image is determined by the sum of the gradient map.

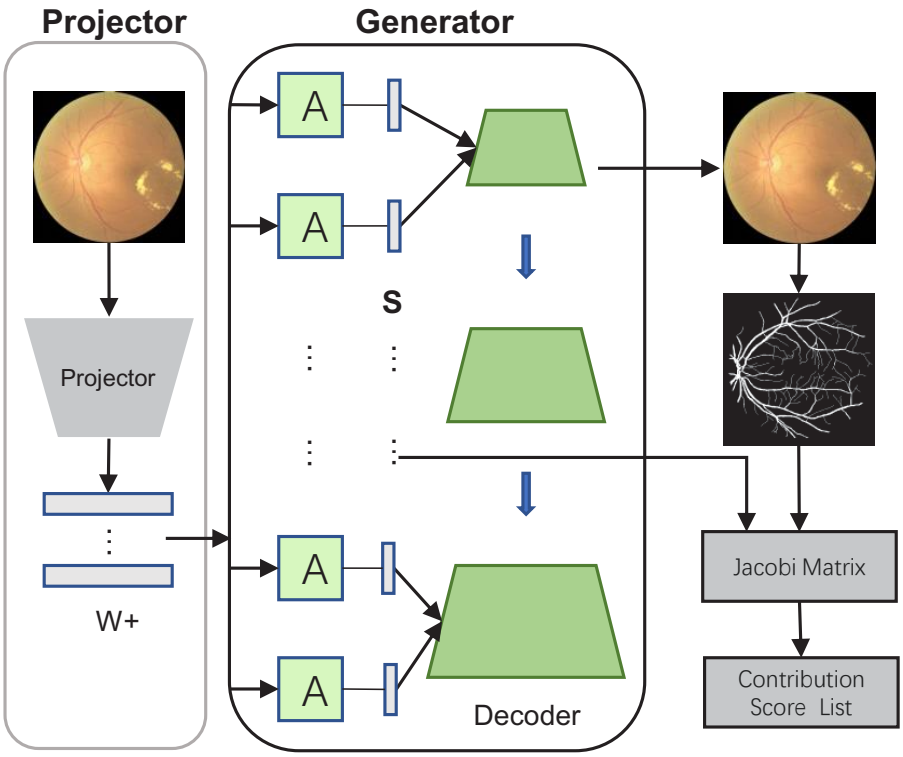

Fig. 2. An overview of proposed structure for locating the latent code with highest contribution score to vascular structure. The structure consists of a projector to invert real fundus image to latent space, a generator to reconstruct the image from latent space. Then we can analyze the contribution score of latent code to vascular structure from the Jacobian matrix (gradient map).

3.2.1. Latent space

Because of the latent space, StyleGAN3 has good disentanglement for different attributes. In the generator structure of StyleGAN3, many types of vectors regarded as latent spaces. For example, a latent code $z \in \mathcal{Z}$ is given as input, the mapping layers $f : \mathcal{Z} \rightarrow \mathcal{W}$, and the affine layers $A : W+ \rightarrow \mathcal{S}$. At first, the latent code $z \in \mathcal{Z}$ is a stochastic normally distributed vector, always called $\mathcal{Z}$ space. After the fully connected mapping layers, z is transformed into a new latent space $\mathcal{W}$. The latent code $w \in \mathcal{W}$ is claimed to have better disentanglement of different attributes [45]. In affine transformation layers of StyleGAN, the input vectors are denoted as $\mathcal{W}+$ space, and the output vectors are denoted as $\mathcal{S}$ space. In [37], the experiments demonstrate that the **style code** $s \in \mathcal{S}$ can better reflect the disentangled nature of the learned distribution than $\mathcal{W}$ and $\mathcal{W}+$ spaces. In order to manipulate real images, it is necessary to invert them into the latent space. There are two main approaches, one is to optimize the latent vector [45,46] and the other is to train an encoder based on reconstruction loss [47,48]. We adapt the latent

Biomedical Optics EXPRESS

optimization algorithm of Karras et al. [27] to invert real images into $\mathcal{W}+$ space. and then mapping the $\mathcal{W}+$ space to $\mathcal{S}$ space.

3.2.2. Jacobian matrix

Image after manipulation x^{edit} can be defined as Eq. (5), which is obtained by moving the latent code in the $\mathcal{S}$ space based on vector $\vec{n}$. Here, $\vec{n}$ means selecting the channel with highest contribution to vascular structure, so we can define $\vec{n}$ as a special unit vector that has only one value of 1 and all others are 0, the length of $\vec{n}$ is the dimensions of S space. α denotes editing strength.

$$x^{\text{edit}} = G(s + \alpha\vec{n}) \tag{5}$$

According to the approach proposed by Zhu et al. [36], we perform a first-order Taylor expansion on Eq. (5). Then, following their notation, the first-order Taylor approximation of x^{edit} performed at point s can be written as:

$$G(s + \alpha\vec{n}) = G(s) + \alpha\mathcal{J}_s \odot \vec{n} + o(\alpha) \tag{6}$$

The Jacobian matrix $\mathcal{J}_s$ is composed of **gradient map** $(\mathcal{J}_s)_k = \frac{(\partial G(s))_k}{(\partial s)_k}$, where k is the $k-th$ dimension of $\mathcal{S}$ space. k can also be mapped as the index of layer and channel (l,c) according to the structure of S space in StyleGAN3. $\mathcal{S}$ space has K dimensions in total. We denote the shape of generated image as $H \times W$, the shape of gradient map of each latent code $(\mathcal{J}_s)_k$ is also $H \times W$, then the shape of $\mathcal{J}_s$ is $K \times H \times W$.

From Eq. (6) The error between original image and edited image can be written as:

$$\varrho = |G(s + \alpha\vec{n}) - G(s)| \leq |\alpha\mathcal{J}_s \odot \vec{n}| + |o(\alpha)| \approx |\alpha\mathcal{J}_s \odot \vec{n}| = \alpha|\mathcal{J}_s| \odot \vec{n} \tag{7}$$

because α is a constant value and $\vec{n}$ means selecting the channel, so we can bring them out of the absolute value operation. In Eq. (7) ϱ denote the error, $|$ denote the absolute value. $o(\alpha)$ is an infinitesimal number. Intuitively, after the manipulation on images, the desired result $G(s)$ should be significantly changed in semantic area. From Eq. (8), as $\vec{n}$ is a special unit vector that represent which channel is selected in $\mathcal{S}$ space, we need to find the channel which has maximum gradient value in the Jacobian matrix of semantic area M. In order to perform vectorized calculations, it is necessary to expand the M into K dimensions as M_K. k is the selected channel, which means the position in $\vec{n}$ with value 1. Here, the operation $\odot$ means element-wise Hadamard product.

$$\arg\max_k \varrho M_K \approx \arg\max_k \alpha|\mathcal{J}_s \odot M_K| \odot \vec{n} = \arg\max_k \alpha|\frac{\partial(G(s) \odot M_K)}{\partial s}| \odot \vec{n}(0<k \leq K) \tag{8}$$

3.2.3. Calculating the contribution score list

After the network model is well-trained, we start to analyze which latent code has the highest contribution score to biometric feature in vascular structure. Firstly, we generate image I^{Gen} of size $H \times W$ from StyleGAN3. For each channel in $\mathcal{S}$ space, the shape of gradient map $(\mathcal{J}_s)_k$ with respect to latent code is $H \times W$, so the Jacobian matrix $\mathcal{J}_s$ of output image with respect to all style code in $\mathcal{S}$ space is $K \times H \times W$. K is defined as the number of channels in $\mathcal{S}$ space. And then we use a segmentation method on output image to get the semantic mask $\boldsymbol{M}$, which has same size of $H \times W$ as I^{Gen}. In order to locate the index of channel with maximum contribution score, we can compute all the gradient maps of latent codes in S space according to the principle of Eq. (8). Since the gradient map is obtained by computing the gradients of the output image with respect to the latent code, based on the vessel mask region, each latent code's gradient map is also an $H \times W$ image. Each gradient in the gradient map represents the contribution of that latent code to the vessel structure pixel in the output image. Therefore, the sum of the entire gradient map of one latent code can be considered as the overall contribution of that latent code to the

Biomedical Optics EXPRESS

entire vessel structure. So, we sum up all the gradient values in the gradient map to represent the contribution score of that latent code to the vessel structure in the output image as shown in Eq. (9). Then we can get a contribution list R, which means the contribution score to vascular structure of each channel. Instead of calculating all pixel values in gradient map, we calculated the top $||M||_0$ values of the largest. Because we only focus on vascular mask, if we count all pixels the result will be affected by region of non-interest. $||M||_0$ is the amount of pixels in mask M. K is the dimensions of S space, and is also the length of n in Eq. (8).

$$R_k = \frac{\sum_{i=1}^{||M||_0} sort(vec(\frac{\partial(I^{\mathrm{Gen}} \odot M)}{\partial_{s_k}}))}{||M||_0} (1 \le i \le ||M||_0) \tag{9}$$

3.3. Vascular structure manipulation

For each style code, we have its contribution score to biometric features on vascular structure. We project T images to style spaces using projector provided by Karras et al. [27], and thereafter, calculate the contribution list for each image. To ensure that the collected positions of style codes are not impacted by individual noise and are consistent across all images, we calculate the mean of all images as shown in Eq. (10). The number of images used for projection is represented by the symbol T, the value of T is specifically defined as 1000 in the paper.

$$\begin{cases} k = arg\max_k(\frac{1}{T}\sum_{i=0}^{T} R^i)_k (0<k \le K) \\ k \Rightarrow (l, c) \end{cases} \tag{10}$$

According to Eq. (10), we need find the latent code in position k , which has highest contribution score. So the contribution list can be sorted from high to low, and then the latent code with highest contribution score is selected as manipulation target. To make it easier to understand k can be mapped as the **index of layer and channel (l, c)** based on the layers and channels of S space in StyleGAN3 network structure.

After finding the general position of latent code that has highest contribution score to vascular structure, we can perform manipulation on real image as shown in Fig. 3 based on Eq. (5). We can project the real image to style space using the projector, then change the value of latent code in position (l, c). However, if we directly add the editing strength α on s, the maximum and minimum value in S space will be changed. In this way, it will make the S space inconsistent with pre-trained StyleGAN3. Therefore, we modify the Eq. (5) to Eq. (11).

$$\begin{cases} S_{new} = s \odot (\mathbf{1} - \vec{n}) + [(1-\alpha)s \odot \vec{n} + \alpha\mu^m \odot \vec{n}] \\ \qquad\quad = s - \alpha(\mu^m - s) \odot \vec{n} \\ I^{de-ID} = G(S_{new}) \end{cases} \tag{11}$$

where $\boldsymbol{I^{de-ID}}$ denotes the image after manipulation, which will be without original identification. $\boldsymbol{\alpha}$ is the editing strength, which is limited to $[0 \sim 1]$. The operation $\odot$ means element-wise Hadamard product. $\mathbf{1}$ is a vector whose values are all 1. μ^m is the mean of style space in ground-truth data distribution, which is from the trained $\boldsymbol{G}$. $\vec{n}$ is a special unit vector that has value 1 at position (l, c), and all others are 0.

3.4. Evaluation metrics

In our experiments, we employ Frechet Inception Distance (FID) [49], Learned Perceptual Image Patch Similarity (LPIPS) [28] and Sliced Wasserstein Difference (SWD) [50]to measure the distance between synthetic and real data distribution. Accuracy, Precision, Recall and F1 score

Biomedical Optics EXPRESS

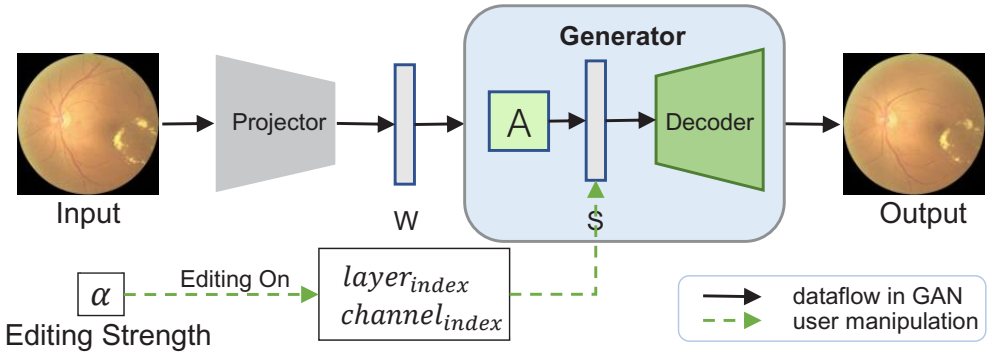

Fig. 3. Manipulation on real image for de-identification on the index of layer and channel with highest contribution score. The difference between output and input image can be controlled by editing strength α.

are applied for human perception score. FS_{SIFT}, FS_{SR}, RROI and AMS metrics are used to measure the similarity between original images and de-identification images. FS_{SIFT}, FS_{SR} focus on the ID similarity by detecting the features and vascular bifurcations. RROI and AMS focus on lesion similarity by calculating the change in pathology region.

AMS Attributes Matching Score (AMS) [51] focus on the whether we change the region of pathologies. AMS use a binary classifier pretrained on hard exudates detection to decide the labels of image pairs, if the labels are the same, we regard them as matching. In our experiments, large AMS score means image pairs still keep same lesion before and after modifying.

RROI RROI compute the ratio of the distance between original and edited images in the region of interest, over the same quantity with the region of 'disinterest'(call it RROI [52]). Concretely, where $\mathcal{M} \in [0, 1]$ $H \times W \times C$ is an $H \times W$ semantic mask of the region of interest, $\mathbf{1}$ is a 1 dimension tensor, and $\mathcal{X}, \mathcal{X}' \in \mathbb{R}^{N \times \tilde{H} \times \tilde{W} \times \tilde{C}}$ are the batch of original and edited versions of the images respectively. A larger RROI indicates a smaller degree of alteration for non-vascular regions, which in this paper means that more pathological information is preserved.

$$\mathrm{RROI}(\mathcal{M}, \mathcal{X}, \mathcal{X}') = \frac{1}{N} \sum_{i=1}^{N} \frac{||\mathcal{M} * (\mathcal{X}_i - \mathcal{X}_i')||}{||(\mathbf{1} - \mathcal{M}) * (\mathcal{X}_i - \mathcal{X}_i')||} \tag{12}$$

FID, LPIPS and SWD The Fréchet inception distance (FID) is a metric used to assess the quality of images created by a generative model. The FID compares the distribution of generated images with the distribution of a set of real images ("ground truth"). The Learned Perceptual Image Patch Similarity (LPIPS) compute the average distance between each generated image and all images in the test set. LPIPS essentially computes the similarity between the activations of two image patches for some pre-defined network. This measure has been shown to match human perception well. Sliced Wasserstein Difference (SWD) aims to capture the natural notion of dissimilarity between the outputs of task-specific classifiers, and SWD directly measures the degree of support between the target sample and the source sample.

FS_{SIFT} and FS_{SR} SIFT [53] Feature Similarity($\boldsymbol{FS_{SIFT}}$) is effective in identifying similarities between images, which can detect, describe, and match local features in images. It performs admirably in image recognition applications. The matching score means the numbers of good match in all detected features. SuperRetina [54] Feature Similarity ($\boldsymbol{FS_{SR}}$) use SuperRetina network to identify the similarity between images. SuperRetina is a network of keypoints detector and descriptor for retinal image matching, which focus on specific vascular points such as crossover and bifurcation that are more stable and repeatable. We perform registration on pair of

Biomedical Optics EXPRESS

images based on the results of keypoints detection. The matching score means the similarity of pair of image.

Precision, Recall, Accuracy, F1 To measure the performance of human perception score, we apply precision, recall, accuracy and F1 score. Precision measures the number of positive predictions that truly belong to the positive class, whereas recall quantifies the number of positive class predictions made out of all positive examples in the dataset. Accuracy represents the number of correctly classified data instances over the total number of data instances. F1 Score is needed when you want to seek a balance between Precision and Recall. As shown in Eq. (13), True positive *TP* is defined as correctly classified data, whereas true negative *TN* is defined as correctly rejected data. Similarly false positive *FP* denotes the incorrectly predicted data and false negative *FN* denotes the incorrectly rejected data.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Dice / F1} &= \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned} \tag{13}$$

4. Experiments

4.1. Dataset

We carry out our experiments on a publicly available dataset named ODIR-5k [55], which is collected by Shanggong Medical Technology Co., Ltd. from different hospitals/medical centers in China. In these institutions, fundus images are captured by various cameras in the market, such as Canon, Zeiss and Kowa, resulting into varied image resolutions (width: [250-5184],height :[188-3456]). The original dataset has 8000 images. There are 1000 images in the testing dataset and 7000 images in the training dataset. Due of inconsistent illumination, several images are of low quality. To overcome this, we use EyeQ [56] to grade the quality of fundus images, which divides them into three levels (Good, Usable, and Reject). Using this strategy, we choose 3000 images from the training set and 500 images from the testing set that have been classified as 'Good' or 'Usable' for our experiments.

4.2. Implementation details

4.2.1. Data preprocessing

The images in ODIR-5k dataset are gathered from multiple clinics using a variety of cameras over an extended period of time so they have different resolutions. Before training, we need to ensure that the input images have the same resolution. To do this, we first detect the edge of the fundus in the entire image and then apply center crop preprocessing method based on the center of the edge. After center crop, we remove the black areas of the image that do not contain any information. We also employ online data augmentation approaches, such as left-right flipping, randomly rotating the image within a range of -30° to 30°, to improve the generalization ability of image generation and prevent overfitting. During each training period, we use one of these augmentation approach randomly.

4.2.2. Experiment setting

In the paper, all the experiments are conducted on a single NVIDIA RTX A5000 GPUs with 24GB memory using PyTorch implementation. We initialize the learning rate with 0.0025 for

Biomedical Optics EXPRESS

$G(\cdot)$ and 0.002 for $D(\cdot)$. For optimization, we use the Adam optimizer with $\beta_1 = 0.0, \beta_2 = 0.99$. We empirically set $\lambda_{lpips} = 0.6, \lambda_{adv} = 0.4$.

4.3. Evaluation and results

4.3.1. Results of synthetic images

For visual evaluation, we show some images generated by StyleGAN3 with our customized LPIPS loss in Fig. 4. The first column of the Fig. 4 shows the fundus images of normal people, while the others show the fundus images of patients of hard exudates. As we can see, synthetic images are visually comparable to real images. In Table 1, we show the quantitative evaluation results. For objective quality evaluation, we employ FID and LPIPS as metrics on 1K testing images to measure the distance between the distribution of synthetic data and real data. For subjective quality evaluation, we conduct a user study to evaluate the results. We randomly chose 200 images which consists of 100 synthetic images and 100 real images, then two professional ophthalmologists in Klinik und Poliklinik für Augenheilkunde, Technische Universität München are asked to determine which images are synthetic. And then we calculate the Accuracy, F1 score, Precision, and Recall. From the results in Table 1, we know that all the evaluation score are close to 0.5. This means that even for professional doctors there is only a 50% possibility to predict which one is real and which one is synthetic. This also implies that synthetic images might be highly confusing even for experienced clinicians.

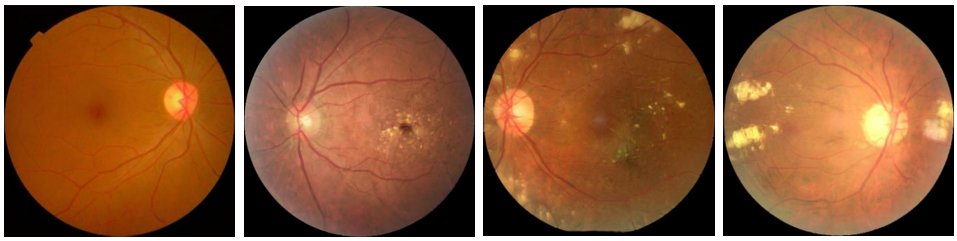

Fig. 4. Generated fundus images by StyleGAN3. The first image is a normal fundus image without any pathology. The second to fourth images are patient image with pathologies.

Table 1. Image Quality Assessment

Objective Quality Metrics		Human Perception Score			
FID(↓)	LPIPS(↓)	Acc	Recall	Precision	F1
9.382	0.457	0.525	0.527	0.480	0.503

4.3.2. Results of de-identification

Due to the good spatial disentangled ability of StyleGAN3 in the latent space, we can control a certain attribute of the generated image by manipulating a specific channel in the latent space. After the training process, we use Jacobian matrix to detect the channel position of style code having the highest contribution to biometric features as described in Section. 3.2.3. Afterward, we observe a list of channels (c) sorted from high to low based on the contribution list. Then we modify the style code in the first position of the list. Our editing results are shown in the third column of Fig. 5. The gradient maps with highest contribution score are shown in the second column. The objective of this paper is to perform de-identification of the image without changing its lesion as much as possible. From the experiments in [26] a well disentanglement latent space, one latent code can control the generation of a specific attribute. Therefore, when editing an

Biomedical Optics EXPRESS

image in latent space, manipulating the whole S space is not necessary. Instead, only the style code with highest contribution score can provide quite good performance.

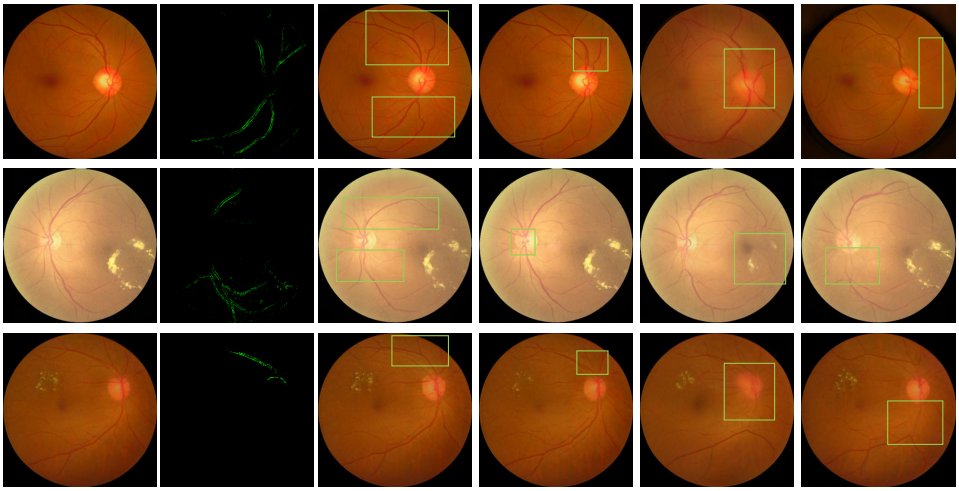

Fig. 5. Comparison results of different methods on image manipulation. The first column is the reference image. The second column is gradient map of latent code with highest contribution score. The third column shows the images after manipulation by our method. The fourth to sixth columns show the results of other manipulation methods.

We also show the comparison with state-of-the-art image local editing methods: LowRank [36], ReSeFa [35] and StyleSpace [37]. The comparison are performed on the same pre-trained StyleGAN3 model with same environment. From the results shown in Fig. 5. All the image manipulation method can modify the target image in a certain direction. StyleSpace always keep the pathologies, when editing the vascular structure, but as shown in third column of Fig. 5, it always change only a small region. LowRank and ReSefa can change more detail in vascular structure, but LowRank will change the lesion and region of interest. ReSefa can keep the region of interest area, but the manipulation of the vascular structure make the vascular structure disappear. If we only anonymous the vascular, it is working well, but the generated data can not be used in some medical image processing tasks, such as vessel segmentation. Our results are shown in the second column, from the images, we can see that our method can edit the vascular structure a lot and keep the pathologies unchanged compared with other methods.

In Fig. 6, we show the visual presentation of FS_{SR} (Feature Similatiry of SuperRetina) [54] on the edited results of LowRank, Resefa, StyleSpace and our method. FS_{SR} can detect the keypoints(cross-over and bifurcation) in fundus images. Based on this property, we can use feature matching methods on the keypoints, then we determine the similarity by calculating the percentage of keypoints that match between the original and edited images, based on the total number of keypoints detected in the original image. As shown in Fig. 6, LowRank can modify a lot on the images, but it almost miss all the keypoints in the lesion area. ReSefa is similar as LowRank in vascular structure, but is much better in lesion area. StyleSpace keeps high similarity except editing region, but the problem is the editing region is always small. If we want to change whole vascular, it is not working well. From our result in Fig. 6, we can see that our method can change the vascular structure as much as possible while keeping lesion as unchanged as possible.

To perform a more comprehensive and quantitative comparison between the proposed method with different local editing approaches, including StyleSpace, LowRank, ReSefa. We set the editing strength α in Eq. (11) as 0.5. Finally, we use FID, LPIPS, FS_{SIFT}, FS_{SR}, RROI and

Biomedical Optics EXPRESS

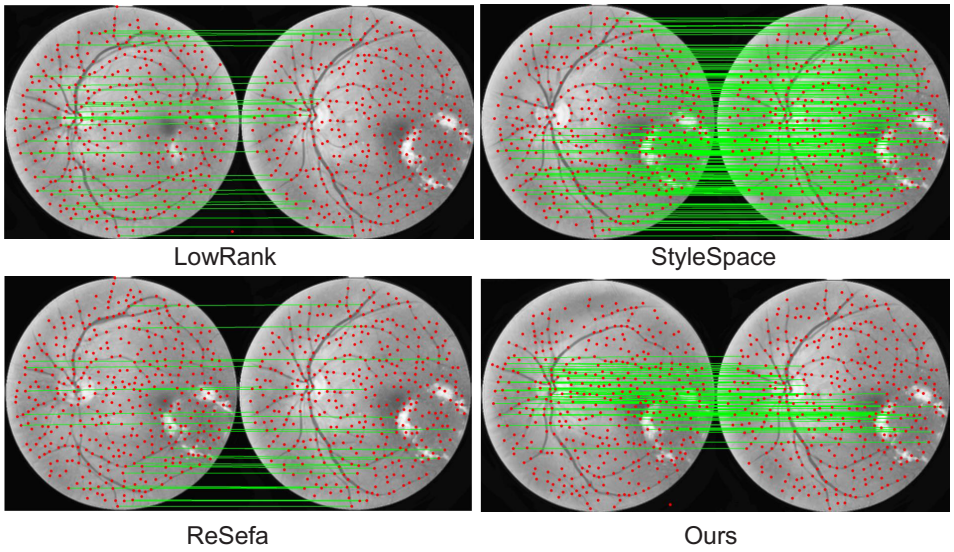

Fig. 6. Visual presentation of FS_{SR} on results of image local editing methods. In each sub-figure the left is the edited image and the right is the reference image. The green lines show the matched bifurcation between original and de-identification image.

AMS metrics on 1K testing images. The quantitative results are shown in Table 2. FID and LPIPS show the distance between generated images and real images. They indicate the realism of synthetic images, the smaller the value, the more realistic it is. After manipulation of target images, the FID and LPIPS score are 26.952 and 0.156 as shown in Table 2, from the FID and LPIPS score in Table 2, we can know our method have the comparable image quality score with StyleSpace, and much better than LowRank and ReSefa. FID and LPIPS results indicate that our method do not changes the data distribution remarkably. We also give the quantitative comparison results of ID similarity. FS_{SIFT} indicates the SIFT feature similarity of an image before and after it has been edited. FS_{SR} indicates the structure similarity which can detect the keypoints of vascular structure in fundus images including crossover and bifurcation. After image manipulation, the ID similarity scores of FS_{SIFT} and FS_{SR} should be as small as possible because there should be a clear distinction in similarity between the images before and after editing. But the preservation of pathologies after editing the vascular structure is as crucial as ID similarity. RROI shows similarity of regions other than the vascular structure, so the smaller the RROI, the better the result. AMS shows whether the original image and the image after editing have different pathological labels, the smaller the AMS, the better the preservation of pathologies.

Table 2. Quantitative comparison with different local editing approaches.

	Image Quality		ID Similarity		Lesion Similarity	
	FID(↓)	LPIPS(↓)	FS_{SIFT}(↓)	FS_{SR}(↓)	RROI(↑)	AMS(↑)
StyleSpace [37]	23.604	0.123	0.676	0.597	0.441	0.990
LowRank [36]	46.486	0.255	0.073	0.206	0.160	0.776
ReSefa [35]	37.542	0.241	0.191	0.274	0.230	0.948
ours	26.952	0.156	0.162	0.271	0.472	0.986

As shown in Table 2, StyleSpace perform very bad in ID similarity, because it always change only a small area, so StyleSpace is not good at de-identification task. LowRank perform well in

Biomedical Optics EXPRESS

de-identification but it change both vascular and lesion. Although ReSefa perform well both in de-identification and lesion keeping, it is not better than our method. And the most important factor in ReSefa is that it makes some vascular structure disappear. The utilization of ReSefa will have a substantial impact when attempting to apply de-identification techniques to the vessel segmentation dataset. The proposed method in this paper outperform ReSefa both in ID similarity and lesion similarity. Compared with StyleSpace and LowRank, our method perform well in a trade-off manner, because ID similarity of the images and the lesion similarity are negatively correlated. Each evaluation of quantitative metrics individually has certain limitations. We show the similarity trade-off score in ablation study section.

4.3.3. Ablation study

LPIPS Loss Functions From left to right columns, we represent the loss function from the training process in Fig. 7. The first column is the original StyleGAN3 results, We can identify the image as a retinal image based on its shape and texture, but the detail processing is quite sloppy, especially in delicate regions like the vascular anatomy. The second column is the result using $LPIPS_{vessel}$ and the third column is the result using $LPIPS_{lesion}$. From the results, we can see that when we add the corresponding LPIPS loss function, the generated image will focus more on these related attributes. Finally, images in last column show that the network with both LPIPS loss perform well in vascular structure and lesion generation. The quantitative evaluation are tabulated in Table 3. Evidently, the proposed loss function based on LPIPS significantly improves the quality of generated images.

Similarity Trade-off Score We have discussed various evaluation metrics for different local editing methods in Table 2. These similarity evaluation methods are mainly divided into two parts: ID similarity and Lesion similarity. In this paper, we talk vascular structure similarity as ID similarity. When we change the vascular structure too much, the lesion will be affected because sometimes vascular structures may pass through the pathological area. So ID similarity and lesion similarity have a negative correlation. In our de-identification experiments, the lower ID similarity score means the better de-identification results. Conversely, the higher Lesion similarity score, the better lesion keeping results. So, when we focus too much on ID similarity, lesion similarity will definitely be greatly impacted, and vice versa. Therefore, we need to find a trade-off in our attention to both similarities. Given that every metric in Table 2 is evaluated on different scales, normalization is essential for obtaining unbiased results. Then we evaluate the impact of different levels of attention on each similarity according to Eq. (14). Here, $ID_{similarity}$ and $Lesion_{similarity}$ are constant values from Table 2.

$$\begin{aligned} \mathrm{ID}_{\mathrm{similarity}} &= \frac{norm(FS_{SIFT}) + norm(FS_{SR})}{2} \\ \mathrm{Lesion}_{\mathrm{similarity}} &= \frac{norm(RROI) + norm(AMS)}{2} \\ \mathrm{SimilarityScore} &= \frac{1.0 - \eta}{\mathrm{ID}_{\mathrm{similarity}}} + \eta \mathrm{Lesion}_{\mathrm{similarity}} \end{aligned} \tag{14}$$

where η represents how much attention we give to Lesion similarity. We can observe from Fig. 8 that Lowrank perform well when the attention to lesion similarity is low and the attention to ID similarity is high, however StyleSpace work badly. Resefa is more steady than stylespace, whose performance is too readily influenced by η. The effectiveness of our method continuously improves as the importance of lesion similarity increases.

Editing Strength Different editing strengths are a trade-off for image quality. The degree of modification for the vascular structure varies when we choose different editing intensity. Due to the widespread distribution of vascular structures throughout the fundus image, they are deeply intertwined with other features in fundus image. Even in the high-dimensional disentangled latent

Biomedical Optics EXPRESS

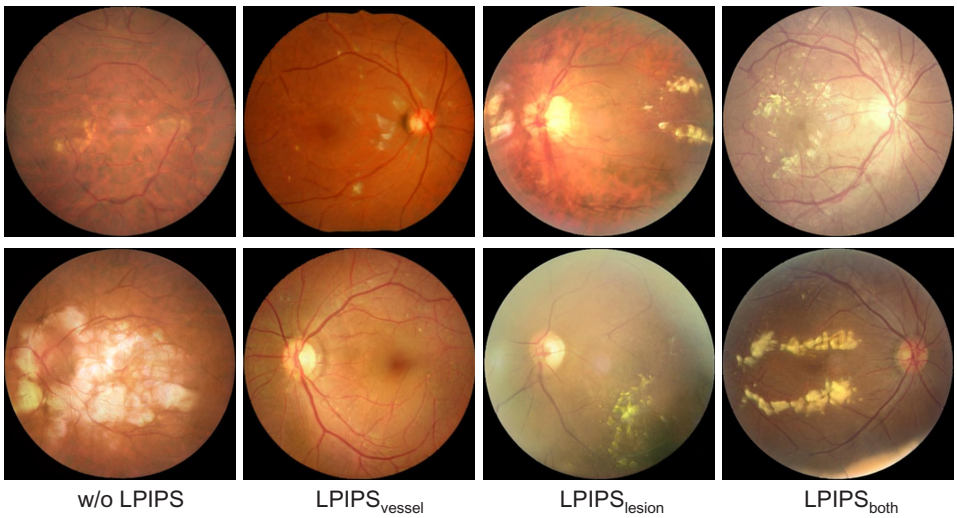

Fig. 7. Visual presentation of LPIPS loss function. The first column show the results of original StyleGAN3. The second column show results with LPIPS loss based on vessel segmentation. The third column show results with LPIPS loss based on lesion segmentation. The last column show the results with both LPIPS loss.

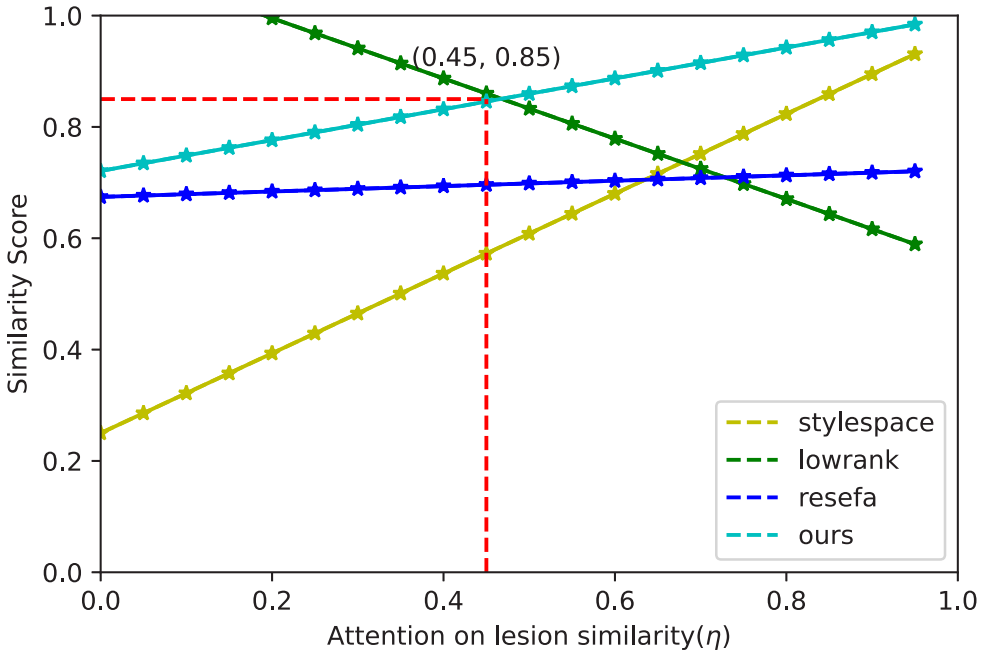

Fig. 8. Similarity Score shows the variation in performance scores of different methods as the attention on the lesion is increased.

Biomedical Optics EXPRESS

Table 3. Comparison on whether to use LPIPS loss

	FID(↓)	LPIPS(↓)	MSE(↓)	SWD(↓)
StyleGAN3	15.279	0.557	0.406	47.109
StyleGAN3+$LPIPS_{vessel}$	12.380	0.540	0.363	43.837
StyleGAN3+$LPIPS_{lesion}$	12.171	0.480	0.347	42.014
StyleGAN3+$LPIPS_{vessel+lesion}$	9.382	0.457	0.283	39.776

space, it is imperative to consider the entanglement between the blood vessels and other features of the fundus image. Particularly, if the vessels traverse pathological regions, the pathological region is more or less affected when we edit the vessel structure. For these reasons, if we choose a high editing intensity, the editing will influence not just the vascular structure but also other areas of the image.

From Fig. 9, we can see that as the magnitude of the editing strength α increases, the changing in the vascular structure becomes more and more obvious, but the lesion is also increasingly affected as well. In the interface stage, we can manually set the editing strength for the vascular structure based on our needs and use similarity calculation and pathological detection methods to determine whether the edited image meets our requirements.

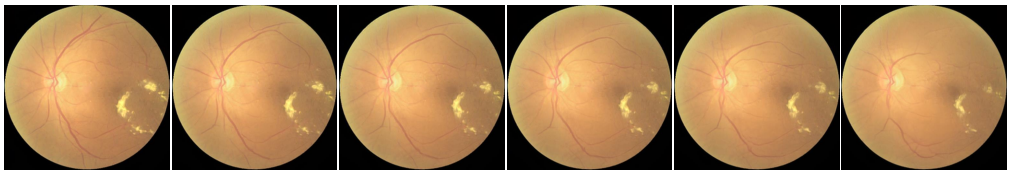

Fig. 9. As α increases, the vascular changes are significant. But when the vessels pass through a lesion, it will also make lesion changed more or less.

As discussed in Sec. 3.4, each evaluation metrics can provide a measurement of image quality, and main focus of each metric is different. As the editing strength α become larger, FID and LPIPS become larger, but this means that the image quality they are focusing on is getting worse. RROI and AMS will become smaller with α become larger, it indicates the image quality focused on by them is becoming worse. FS_{SIFT} and FS_{SR} are also become smaller, but it indicated the ID similarity is becoming better, because the new images after editing are becoming more different from original one. To make every evaluation metric in same scale, we normalize the different metrics between $[0-1]$. In Fig. 10, we show the impact of different editing strength α on various evaluation metrics results. As the editing strength increases, the difference between the edited image and the original image is also increasing. The data distribution of the generated image and the real image are increasing, so the FID and LPIPS are increasing, which means the image realism is getting worse. The difference of lesions between edited images and original images are also increasing, which means the lesions similarity is becoming worse, so the RROI and AMS metrics decrease. However, with the increasing of α, the difference of vessels become larger, the ID similarity decrease, it means the performance of de-identification is improving.

It is important to note that these results are a mean of multiple images and may require further adjustment for individual images to ensure the editing strength α can meet specific requirements.

Biomedical Optics EXPRESS

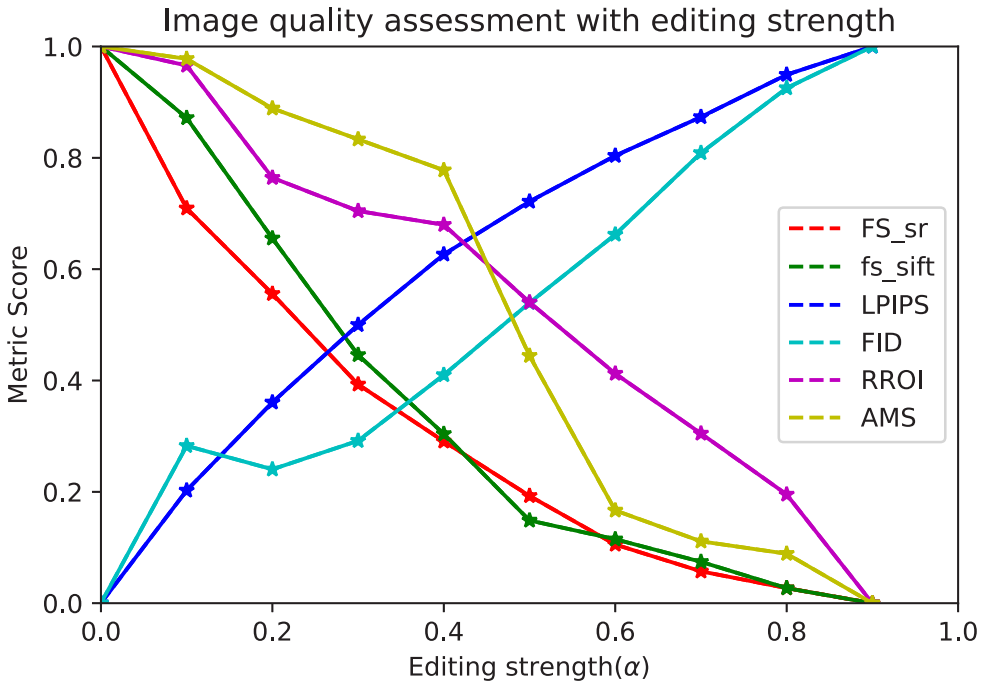

Fig. 10. As the editing strength increases, the increase of FID and LPIPS metric show that image quality is getting worse, the decrease of RROI and AMS show that quality in pathology region is getting worse, the decrease of FS_{SIFT} and FS_{SR} show that the performance of de-identification is getting better.

5. Conclusion

In this paper we propose a de-identification method on vascular structure of fundus images, which can preserve the pathologies when removing the biometric features. In this work, the LPIPS losses based on hard exudates detection and vessel segmentation were adopted into our method. However, the challenge that was not addressed in this part of the study was differentiation between safe and critical vascular segments which might effect the prognosis of disease related with vessels. This challenge will be addressed in our future work together with our collaborating ophthalmologists within a qualitative research scheme.

Funding. Bayerische Forschungsstiftung (AZ-1503-21).

Acknowledgments. The authors would like to thank all the participants for their interests in this research.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Kaggle [55].

References

1. T. Li, W. Bo, C. Hu, H. Kang, H. Liu, K. Wang, and H. Fu, "Applications of deep learning in fundus images: A review," Med. Image Anal. **69**, 101971 (2021).
2. E. Tom, P. A. Keane, M. Blazes, L. R. Pasquale, M. F. Chiang, A. Y. Lee, C. S. Lee, and A. A. I. T. Force, "Protecting data privacy in the age of ai-enabled ophthalmology," Trans. Vis. Sci. Tech. **9**(2), 36 (2020).
3. J. Segal and M. J. Sacopulos, "Photography consent and related legal issues," Facial Plastic Surgery Clinics **18**(2), 237–244 (2010).
4. C. A. Koch and W. F. Larrabee, "Patient privacy, photographs, and publication," JAMA Facial Plast. Surg. **15**(5), 335–336 (2013).
5. M. Jayabalan and M. E. Rana, "Anonymizing healthcare records: a study of privacy preserving data publishing techniques," Adv. Sci. Lett. **24**(3), 1694–1697 (2018).

Biomedical Optics EXPRESS

6. P. Riis and M. Nylenna, "Patients have a right to privacy and anonymity in medical publication," JAMA **265**(20), 2720 (1991).
7. M. Suganya and K. Krishnakumari, "A novel retina based biometric privacy using visual cryptography," International Journal of Computer Science and Network Security (IJCSNS) **16**(9), 76 (2016).
8. Z. Waheed, M. U. Akram, A. Waheed, M. A. Khan, A. Shaukat, and M. Ishaq, "Person identification using vascular and non-vascular retinal features," Comput. & Electr. Eng. **53**, 359–371 (2016).
9. A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Trans. Circuits Syst. Video Technol. **14**(1), 4–20 (2004).
10. F. Jiu, K. Noronha, and D. Jayaswal, "Biometric identification through detection of retinal vasculature," in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, (IEEE, 2016), pp. 1–5.
11. J. Fatima, A. M. Syed, and M. U. Akram, "A secure personal identification system based on human retina," in *2013 IEEE Symposium on Industrial Electronics & Applications*, (IEEE, 2013), pp. 90–95.
12. P. Elangovan and M. K. Nath, "A review: person identification using retinal fundus images," Int. J. Electron. Telecommun. **65**(4), 585–596 (2023).
13. R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), pp. 5485–5493.
14. Z. Chen, T. Zhu, P. Xiong, C. Wang, and W. Ren, "Privacy preservation for image data: a gan-based method," Int. J. Intell. Syst. **36**(4), 1668–1685 (2021).
15. Q. Sun, A. Tewari, W. Xu, M. Fritz, C. Theobalt, and B. Schiele, "A hybrid model for identity obfuscation by face replacement," in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), pp. 553–569.
16. Q. Sun, L. Ma, S. J. Oh, L. Van Gool, B. Schiele, and M. Fritz, "Natural and effective obfuscation by head inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), pp. 5050–5059.
17. A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," Multiscale Model. Simul. **4**(2), 490–530 (2005).
18. S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, (2020), pp. 2437–2445.
19. X. Chen, L.-J. Li, L. Fei-Fei, and A. Gupta, "Iterative visual reasoning beyond convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), pp. 7239–7248.
20. L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou, "Differentially private generative adversarial network," arXiv arXiv:1802.06739 (2018).
21. W. Paul, Y. Cao, M. Zhang, and P. Burlina, "Defending medical image diagnostics against privacy attacks using generative methods: Application to retinal diagnostics," in *Clinical Image-Based Procedures, Distributed and Collaborative Learning, Artificial Intelligence for Combating COVID-19 and Secure and Privacy-Preserving Machine Learning*, (Springer, 2021), pp. 174–187.
22. M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, "Federated learning and differential privacy for medical image analysis," Sci. Rep. **12**(1), 1953 (2022).
23. J. S. Chen, A. S. Coyner, R. P. Chan, M. E. Hartnett, D. M. Moshfeghi, L. A. Owen, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, "Deepfakes in ophthalmology: Applications and realism of synthetic retinal images from generative adversarial networks," Ophthalmol. Sci. **1**(4), 100079 (2021).
24. C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Generative adversarial privacy," arXiv arXiv:1807.05306 (2018).
25. Y. Nitzan, A. Bermano, Y. Li, and D. Cohen-Or, "Face identity disentanglement via latent space mapping," arXiv arXiv:2005.07728 (2020).
26. D. Bau, J.-Y. Zhu, H. Strobelt, B. Zhou, J. B. Tenenbaum, W. T. Freeman, and A. Torralba, "Gan dissection: Visualizing and understanding generative adversarial networks," arXiv arXiv:1811.10597 (2018).
27. T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," in *Proc. NeurIPS*, (2021).
28. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, (2018).
29. E. Poonguzhali, R. Giritharan, M. K. Nath, and O. P. Acharya, "Review on localization of optic disc in retinal fundus images," in *2018 International Conference on Applied Electromagnetics, Signal Processing and Communication (AESPC)*, vol. 1 (IEEE, 2018), pp. 1–7.
30. J. B. Mazumdar and S. Nirmala, "Deep learning framework for biometric authentication using retinal images," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization pp. 1–10 (2022).
31. Y. Alaluf, O. Tov, R. Mokady, R. Gal, and A. Bermano, "Hyperstyle: Stylegan inversion with hypernetworks for real image editing," in *Proceedings of the IEEE/CVF conference on computer Vision and pattern recognition*, (2022), pp. 18511–18521.
32. S. Khodadadeh, S. Ghadar, S. Motiian, W.-A. Lin, L. Bölöni, and R. Kalarot, "Latent to latent: A learned mapper for identity preserving editing of multiple face attributes in stylegan-generated images," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (2022), pp. 3184–3192.

Biomedical Optics EXPRESS

33. X. Yao, A. Newson, Y. Gousseau, and P. Hellier, "Learning non-linear disentangled editing for stylegan," in *2021 IEEE International Conference on Image Processing (ICIP)*, (IEEE, 2021), pp. 2418–2422.
34. Y. Shen, C. Yang, X. Tang, and B. Zhou, "Interfacegan: Interpreting the disentangled face representation learned by gans," TPAMI (2020).
35. J. Zhu, Y. Shen, Y. Xu, D. Zhao, and Q. Chen, "Region-based semantic factorization in gans," arXiv arXiv:2202.09649 (2022).
36. J. Zhu, R. Feng, Y. Shen, D. Zhao, Z.-J. Zha, J. Zhou, and Q. Chen, "Low-rank subspaces in gans," Advances in Neural Information Processing Systems **34**, 16648–16658 (2021).
37. Z. Wu, D. Lischinski, and E. Shechtman, "Stylespace analysis: Disentangled controls for stylegan image generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), pp. 12863–12872.
38. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*, (Springer, 2016), pp. 694–711.
39. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv arXiv:1409.1556 (2014).
40. O. Russakovsky, J. Deng, and H. Su, *et al.*, "Imagenet large scale visual recognition challenge," Int. J. Comp. Vis. **115**(3), 211–252 (2015).
41. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
42. C. Eastwood and C. K. Williams, "A framework for the quantitative evaluation of disentangled representations," in *International Conference on Learning Representations*, (2018).
43. K. Ridgeway and M. C. Mozer, "Learning deep disentangled embeddings with the f-statistic loss," Advances in neural information processing systems **31** (2018).
44. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, (2017), pp. 618–626.
45. R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan: How to embed images into the stylegan latent space," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), pp. 4432–4441.
46. R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan++: How to edit the embedded images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2020), pp. 8296–8305.
47. J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*, (Springer, 2016), pp. 597–613.
48. J. Zhu, Y. Shen, D. Zhao, and B. Zhou, "In-domain gan inversion for real image editing," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16*, (Springer, 2020), pp. 592–608.
49. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," Advances in neural information processing systems **30** (2017).
50. C.-Y. Lee, T. Batra, M. H. Baig, and D. Ulbricht, "Sliced wasserstein discrepancy for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2019), pp. 10285–10295.
51. M. J. Chong, W.-S. Chu, A. Kumar, and D. Forsyth, "Retrieve in style: Unsupervised facial feature transfer and retrieval," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), pp. 3887–3896.
52. J. Oldfield, C. Tzelepis, Y. Panagakis, M. A. Nicolaou, and I. Patras, "Panda: Unsupervised learning of parts and appearances in the feature maps of gans," arXiv arXiv:2206.00048 (2022).
53. P. C. Ng and S. Henikoff, "Sift: Predicting amino acid changes that affect protein function," Nucleic Acids Res. **31**(13), 3812–3814 (2003).
54. J. Liu, X. Li, Q. Wei, J. Xu, and D. Ding, "Semi-supervised keypoint detector and descriptor for retinal image matching," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXI*, (Springer, 2022), pp. 593–609.
55. S. M. Technology, "Ocular disease recognition," Kaggle, 2020, https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k.
56. H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, and L. Shao, "Evaluation of retinal image quality assessment networks in different color-spaces," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer, 2019), pp. 48–56.